ECE 636: Systems identification
Lectures 13-14

Input signals
Model parametrization
• Linear regression – Model selection:
• For GWN $F = \frac{(V_1 - V_2) / (d_1 - d_2)}{V_2 / (N - d_2)}$

$$F < \chi^2_{d_2 - d_1, \alpha} : M1 \quad F > \chi^2_{d_2 - d_1, \alpha} : M2$$

follows a $F_{d_2 - d_1, N - d_2}$ (approx. $\chi^2_{d_2 - d_1, \alpha}$ for large N)

• Numerical issues, condition number of $\Phi$ or $\Phi^T \Phi$ (dependence on input characteristics for systems ID problems)
• QR decomposition
• Singular value decomposition
• Regularization

$$W_N(\Theta) = \frac{1}{2} \sum_{k=1}^{N} [y_k - g(\phi_k)]^2 + \frac{\lambda}{2} \sum_{j=1}^{d} |\theta_j|^q$$

$q = 0.5 \quad q = 1 \quad q = 2 \quad q = 4$
- Linear regression and linear systems identification
- Impulse response models

\[ y(t) = \sum_{\tau=0}^{M-1} h(\tau)u(t-\tau) + e(t) \]

\[ \theta = [h(0) \quad h(1) \quad \ldots \quad h(M-1)]^T \]

\[ \varphi(t) = [u(t) \quad u(t-1) \quad \ldots \quad u(t-M+1)] \]

\[ \hat{\theta}_{LS} = \begin{bmatrix}
\frac{1}{N} \sum_{i=1}^{N} y^2(t) & \frac{1}{N} \sum_{i=1}^{N} y(t-1)u(t-1) & \cdots & \frac{1}{N} \sum_{i=1}^{N} y(t-M+1)u(t-M+1) \\
\frac{1}{N} \sum_{i=1}^{N} u(t-1) & \frac{1}{N} \sum_{i=1}^{N} y^2(t-1) & \cdots & \frac{1}{N} \sum_{i=1}^{N} y(t-1)u(t-M+1) \\
\vdots & \vdots & \ddots & \vdots \\
\frac{1}{N} \sum_{i=1}^{N} u(t-M+1) & \frac{1}{N} \sum_{i=1}^{N} u(t-M+1)u(t-1) & \cdots & \frac{1}{N} \sum_{i=1}^{N} y^2(t-M+1) \\
\end{bmatrix}^{-1}
\begin{bmatrix}
\frac{1}{N} \sum_{i=1}^{N} y(t)y(t) \\
\frac{1}{N} \sum_{i=1}^{N} y(t-1)y(t) \\
\vdots \\
\frac{1}{N} \sum_{i=1}^{N} y(t-M+1)y(t) \\
\end{bmatrix} \]

- ARX models

\[ y(t) = \varphi^T(t)\theta + e(t) \]

\[ \theta = [a_1 \quad \ldots \quad a_{n_a} \quad b_1 \quad \ldots \quad b_{n_b}]^T \]

\[ \varphi(t) = [-y(t-1) \quad \ldots \quad -y(t-n_a) \quad u(t-1) \quad \ldots \quad u(t-n_b)]^T \]

\[ \hat{\theta}_{LS} = \begin{bmatrix}
\frac{1}{N} \sum_{i=1}^{N} y^2(t-1) & \frac{1}{N} \sum_{i=1}^{N} y(t-1)y(t-2) & \cdots & -\frac{1}{N} \sum_{i=1}^{N} y(t-1)u(t-n_b) \\
\frac{1}{N} \sum_{i=1}^{N} y(t-1)y(t-2) & \frac{1}{N} \sum_{i=1}^{N} y^2(t-1) & \cdots & -\frac{1}{N} \sum_{i=1}^{N} y(t-2)u(t-n_b) \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{1}{N} \sum_{i=1}^{N} y(t-n_b)y(t-1) & -\frac{1}{N} \sum_{i=1}^{N} y(t-n_b)y(t-2) & \cdots & \frac{1}{N} \sum_{i=1}^{N} y^2(t-n_b) \\
\end{bmatrix}^{-1}
\begin{bmatrix}
-\frac{1}{N} \sum_{i=1}^{N} y(t-1)y(t) \\
-\frac{1}{N} \sum_{i=1}^{N} y(t-2)y(t) \\
\vdots \\
\frac{1}{N} \sum_{i=1}^{N} y(t-n_b)y(t) \\
\end{bmatrix} \]
• Maximum likelihood estimation  \( \hat{\theta}_{ML} = \arg \max_{\theta} p(y \mid \theta) \)

• Example: mean and variance of normally distributed univariate i.i.d. observations

\[
\hat{\mu}_{ML} = \frac{1}{N} \sum_{i=1}^{N} y_i \quad \hat{\sigma}^2_{ML} = \frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{\mu}_{ML})^2
\]

• Example: mean and covariance matrix of multivariate normally distributed i.i.d. observations

\[
\hat{\mu}_{ML} = \frac{1}{N} \sum_{i=1}^{N} x_i \quad \Sigma_{ML} = \frac{1}{N} \sum_{n=1}^{N} (x_n - \mu_{ML})(x_n - \mu_{ML})^T.
\]

• For linear regression models and GWN  \( y(t) = \phi^T(t)\theta + e(t) \quad e \sim N(0, \lambda^2 I) \)

\[
\ln p(y \mid \theta, \lambda^2) = \sum_{t=1}^{N} \ln N(\phi^T(t)\theta, \lambda^2) = -\frac{1}{2\lambda^2} \sum_{t=1}^{N} (y(t) - \phi^T(t)\theta)^2 - \frac{N}{2} \ln(\lambda^2) - \frac{N}{2} \ln(2\pi)
\]

the ML estimate is identical to the LS estimate

\[
\hat{\theta}_{ML} = \hat{\theta}_{LS} = (\Phi^T\Phi)^{-1}\Phi^Ty
\]

and  \( \hat{\lambda}_{ML} = \frac{1}{N} \sum_{t=1}^{N} (y(t) - \phi^T(t)\hat{\theta}_{ML})^2 \)

• Maximum a posteriori estimation

\[
p(\theta \mid y) = \frac{p(y, \theta)}{p(y)} = \frac{p(y \mid \theta)p(\theta)}{p(y)} \Rightarrow \text{Posterior} = \frac{\text{Likelihood} \times \text{Prior}}{\text{Normalizing Constant}}
\]

\[
\hat{\theta}_{MAP} = \arg \max_{\theta} p(\theta \mid y) = \arg \max_{\theta} p(y \mid \theta)p(\theta)
\]
• \( p(\theta) \): prior distribution
• Example: linear regression with Gaussian zero mean prior 
  \[
p(\theta) = N(0, \alpha^2 I) = \left( \frac{1}{2\pi\alpha} \right)^{d/2} \exp\left( -\frac{1}{2\alpha^2} \|\theta\|^2 \right)
\]

\[
\ln p(y|\theta) + \ln p(\theta) = -\frac{1}{2\sigma^2} \sum_{t=1}^{N} (y(t) - \phi^T(t)\theta)^2 - \frac{N}{2} \ln(\lambda^2) - \frac{N}{2} \ln(2\pi) - \frac{1}{2\alpha^2} \sum_{i=1}^{d} w_i^2 - \frac{d}{2} \ln(\alpha^2) - \frac{d}{2} \ln(2\pi)
\]

• Equivalent to minimizing

\[
\frac{1}{2\lambda^2} \sum_{t=1}^{N} (y(t) - \phi^T(t)\theta)^2 + \frac{1}{2\alpha^2} \sum_{i=1}^{d} \theta_i^2
\]

(regularization)

• Bayesian estimation – compute the posterior distribution \( p(\theta|y) \)
Input signals

• We have already seen (Lectures 7-8) that the input type may influence the results systems identification very significantly

• Basic input signals used in practice:
  • Step signal
  • Pseudorandom binary sequences
  • ARMA processes
  • Sinusoidal sums
  • Spontaneous signal fluctuations  (e.g. physiological signals often exhibit broadband characteristics)

• Step signal:
  • More useful for estimating pure time delays and time constants
Input signals

- Note that for a **deterministic signal** we can define its mean and autocorrelation function as:
  \[
  \mu = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} y(t)
  \]
  \[
  \phi_{xx}(\tau) = \lim_{N \to \infty} \frac{1}{N} \sum_{t=1}^{N} [y(t)y(t + \tau)]
  \]
  assuming the corresponding limits exist. These definitions are as we saw equivalent to the standard stochastic definitions:
  \[
  \mu = E[y(t)]
  \]
  \[
  \phi_{xx}(\tau) = E[y(t)y(t + \tau)]
  \]
  for **ergodic** stochastic processes
- Also the spectrum of a deterministic signal is defined, similarly to the case of stochastic signals, as the FT of its autocorrelation function, i.e. we have the following FT pair:
  \[
  \Phi_{xx}(\omega) = \sum_{\tau=-\infty}^{\infty} \phi_{xx}(\tau)e^{-j\omega\tau}
  \]
  \[
  \phi_{xx}(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Phi_{xx}(\omega)e^{j\omega\tau} \, d\omega
  \]
Input signals

• Pseudorandom binary sequences
  • Two possible values, alternation between them is specific
  • These sequences approach the properties of white noise and may be implemented in practice
  • Periodic signal, typically its period $M$ is equal or larger than the number of samples $N$ in an experiment
  • For a sequence alternating between $\alpha$ and $-\alpha$ with a period $M$ (proof Soderstrom Complement C5.3):

\[
\varphi_{\text{int}}(\tau) = \begin{cases} 
  a^2, & \tau = 0, \pm M, \pm 2M, \ldots \\
  -a^2 / M, & \text{otherwise}
\end{cases}
\]

\[
\Phi_{\text{int}}(\omega) = \frac{a^2}{M^2} \left[ \delta(\omega) + (M + 1) \sum_{k=1}^{M-1} \delta(\omega - 2\pi \frac{k}{M}) \right]
\]

The larger $M$ becomes, the more we approach white noise ($\phi$ approaches a delta function and the delta functions in $\Phi$ are more closely placed to each other).

• Note that while the second-order (covariance functions) of PRBS resemble those of Gaussian white noise, its distribution function does not! (obvious, since PRBS is binary)
Pseudorandom binary sequences

- PRBS can be implemented by shift registers of order $n$
- All state variables are binary (0 or 1)
- At each clock pulse (equal to the sampling frequency), the value of the $k$-th state is transferred to the $(k+1)$-th state and the output is taken by the last state variable
- $\oplus$ denotes modulo-2 addition (OR operation)
- The feedback coefficients are also 0 or 1
- PRBS is a purely deterministic signal, however its autocorrelation sequence resembles that of white noise
- For $n$ states we have $2^n$ possible state vectors and the maximum period for a PRBS is $2^n-1$. A PRBS with this period is called maximum length PRBS. Whether a max length PRBS is created depends on the design of the shift register (i.e., on the feedback coefficients), e.g. for $n=3$, $a_1=1$, $a_2=1$, $a_3=0$ and initial state $(1 \ 0 \ 0)^T$ we have the sequence $(1 \ 0 \ 0)^T$, $(1 \ 1 \ 0)^T$, $(0 \ 1 \ 1)^T$, $(1 \ 0 \ 1)^T$, $(1 \ 1 \ 0)^T$ so we have a period of three. For $a_1=1$, $a_2=0$, $a_3=1$ we have the max period (=7).
- Max length PRBS is desirable as it resembles white noise more (recall that $\varphi_{ww}(\tau) = \begin{cases} a_1^2, & \tau = 0, \pm M, \pm 2M, \ldots \\ -a_1^2 / M, & \text{otherwise} \end{cases}$)
- Design of max length PRBS – Complement C5.3
Pseudorandom binary sequences

- Example: Filtered GWN and PRBS
- Let the following two signals
  
  \[ y_1(t) - ay_1(t - 1) = u_1(t) \]
  
  \[ y_2(t) - ay_2(t - 1) = u_2(t) \]

  where \( u_1(t) \) is GWN \( N(0, \lambda^2) \) and \( u_2(t) \) PRBS. Covariance function of \( y_1(t) \):

  \[
  \phi_{y_1y_1}(\tau) = \frac{\lambda^2}{1 - a^2} |a|^{-|\tau|}
  \]

  For \( y_2(t) \):

  \[
  \Phi_{y_2y_2}(\omega) = \Phi_{u_3u_3}(\omega)|H(\omega)|^2
  \]

  \[
  = \frac{\lambda^2}{M^2} \left| \frac{1}{1 - ae^{-j\omega}} \right|^2 \left[ \delta(\omega) + (M + 1) \sum_{k=1}^{M-1} \delta(\omega - 2\pi \frac{k}{M}) \right]
  \]

  Therefore:

  \[
  \phi_{y_2y_2}(\tau) = \Re \int_0^{2\pi} \Phi_{y_2y_2}(\omega)e^{j\tau \omega} d\omega
  \]

  \[
  = \int_0^{2\pi} \frac{\lambda^2}{M^2} \left| \frac{1}{1 - ae^{-j\omega}} \right|^2 \left[ \delta(\omega) + (M + 1) \sum_{k=1}^{M-1} \delta(\omega - 2\pi \frac{k}{M}) \right] \cos(\tau\omega) d\omega
  \]

  \[
  = \frac{\lambda^2}{M^2} \left[ \frac{1}{1 - a^2} + (M + 1) \sum_{k=1}^{M-1} \frac{1}{1 + a^2 - 2a \cos(2\pi k/M)} \cos(\tau 2\pi \frac{k}{M}) \right]
  \]
Pseudorandom binary sequences

• For larger $M$, the two sequences resemble each other more

\[ \phi_{y_2y_2}(\tau) = \frac{\lambda^2}{M^2} \left[ \frac{1}{1 - a^2} + (M + 1) \sum_{k=1}^{M-1} \frac{1}{1 + a^2 - 2a \cos(2\pi k/M)} \cos(\tau 2\pi \frac{k}{M}) \right] \]

FIGURE 5.6 Plots of $r_1(\tau)$ and $r_2(\tau)$ versus $\tau$. The filter parameter is $a = 0.9$; $r_2(\tau)$ is plotted for $M = 500, 200, 100, 50$ and $20$. 
Input signals

- **ARMA processes**
  
  \[ u(t) + a_1 u(t-1) + \ldots + a_{n_a} u(t-n_a) = e(t) + c_1 e(t-1) + \ldots + c_{n_c} e(t-n_c) \]
  
  \[ A(q^{-1})u(t) = C(q^{-1})e(t) \]

  Where \( e(t) \) GWN \( N(0, \lambda^2) \) – \( u(t) \) is essentially filtered white noise

- Depending on the selected coefficients and consequently the filter characteristics, we can obtain a variety of spectral characteristics for \( u(t) \)

- Spectrum:
  
  \[ \Phi_{uu}(\omega) = \lambda^2 \left| \frac{C(\omega)}{A(\omega)} \right|^2 \]

  - If the zeros of \( A(\omega) \) lie near the unit circle (e.g. at \( e^{\pm j\omega_0} \)) -> the spectrum of \( u \) exhibits peaks close to the resonance frequency \( \omega_0 \)
  - Similarly, if \( B(\omega) \) has zeros close to \( e^{\pm j\omega_0} \) -> the spectrum components of \( u \) close to \( \omega_0 \) are negligible

Example: ARMA(2,2)
Input signals

• Sum of sinusoidal signals
  \[ u(t) = \sum_{j=1}^{m} a_j \sin(\omega_j t + \varphi_j) \]

Choice of  \( a_j, \omega_j, \varphi_j \)
• Autocorrelation sequence (proof – Soderstrom Example 5.7)
  \[ \varphi_{uu}(\tau) = \sum_{j=1}^{m} \frac{a_j^2}{2} \cos(\omega_j \tau + \varphi_j) \]

• Spectrum \( \Phi_{uu}(\omega) = \sum_{j=1}^{m} \frac{a_j^2}{4} [\delta(\omega - \omega_j) + \delta(\omega + \omega_j)] \)

Ex: Sum of 4 sinusoids
Input signals

- Sum of sinusoidal signals

A sum of two sinusoids ($a_1 = 1$, $a_2 = 2$, $\omega_1 = 0.4$, $\omega_2 = 0.7$, $\varphi_1 = \varphi_2 = 0$).
Persistent excitation

- **Definition**: A signal \( u(t) \) is termed persistently exciting of order \( n \) if:
  - The limit \( \varphi_{uu}(\tau) = \frac{1}{N} \lim_{N \to \infty} \sum_{t=1}^{N} u(t+\tau)u(t) \) exists
  - The matrix
    \[
    \Phi_{uu}(n) = \begin{bmatrix}
    \varphi_{uu}(0) & \varphi_{uu}(1) & \cdots & \varphi_{uu}(n-1) \\
    \varphi_{uu}(1) & \varphi_{uu}(0) & \cdots & \varphi_{uu}(n-2) \\
    \cdots & \cdots & \cdots & \cdots \\
    \varphi_{uu}(1-n) & \varphi_{uu}(2-n) & \cdots & \varphi_{uu}(0)
    \end{bmatrix}
    \]
  is positive definite. This is equivalent to that its determinant is positive and all the 1x1, 2x2 etc principal minors (upper left sub-matrices) have positive determinants as \( \Phi_{uu} \) is additionally a symmetric matrix.
  For a zero-mean ergodic process, this matrix is equal to the cross-covariance matrix, as the limits are asymptotically equal to the expected value operator \( \mathbb{E}\{.\} \)
- **Reminder**: We saw that when applying correlation analysis, we had:
  \[
  \begin{bmatrix}
  \hat{\varphi}_{uy}(0) \\
  \hat{\varphi}_{uy}(1) \\
  \cdots \\
  \hat{\varphi}_{uy}(M-1)
  \end{bmatrix} = \begin{bmatrix}
  \hat{\varphi}_{uu}(0) & \hat{\varphi}_{uu}(-1) & \cdots & \hat{\varphi}_{uu}(-(M-1)) \\
  \hat{\varphi}_{uu}(1) & \hat{\varphi}_{uu}(0) & \cdots & \hat{\varphi}_{uu}(-(M-2)) \\
  \cdots & \cdots & \cdots & \cdots \\
  \hat{\varphi}_{uu}(M-1) & \hat{\varphi}_{uu}(M-2) & \cdots & \hat{\varphi}_{uu}(0)
  \end{bmatrix}\begin{bmatrix}
  \hat{h}(0) \\
  \hat{h}(1) \\
  \cdots \\
  \hat{h}(M-1)
  \end{bmatrix} = \left(\Phi_{uu}\right)^{-1} \Phi_{uy}
  \]
  Therefore, in order to obtain a unique solution the matrix \( \Phi_{uu} \) must be non-singular, in other words the input should be persistently exciting of order at least \( M \) (where \( M \) is the system memory).
Persistent excitation

• Examples:
• Gaussian white noise is p.e. of any order n as:
  \( \Phi_{uu}(n) = \sigma^2 I_n \) -> positive definite for any n
• The step signal with amplitude \( \sigma \) is p.e. of order n=1 only as
  \( \varphi_{uu}(\tau) = \sigma^2 \Rightarrow \Phi_{uu}(n) \) nonsingular only for n=1
• The impulse signal is not p.e. of any order as
  \( \varphi_{uu}(\tau) = 0 \Rightarrow \Phi_{uu}(n) \) singular for any n

• For noise-free systems, persistent excitation is not a necessary condition for accurate estimation, e.g. if we don’t have noise even the impulse response method will yield accurate estimates for a finite number of samples. However, for noisy systems if we use e.g. correlation analysis for a system of memory M or the least-squares method for a linear system of order n, the input has to be p.e. of order M or n respectively in order to obtain consistent estimates for \( N \to \infty \)
• The condition number of the matrix \( \Phi_{uu}(n) \) may be also used as a measure of the persistency properties of an input signal.
• A signal \( u(t) \) that is p.e. of order n has a spectral density that is nonzero in at least n frequencies.
• If \( u(t) \) is p.e. of order n and \( H(q^{-1}) \) is a stable filter with k zeros on the unit circle, then its output \( y(t)=H(q^{-1})u(t) \) is p.e. of order m where n-k≤m≤n. If \( H(q^{-1}) \) has no zeros on the unit circle, \( u(t) \) and \( y(t) \) will be p.e. of the same order.
• A general periodic signal of period M can be p.e. of order at most M
• A sum of m sinusoids signal can be p.e. of order at most 2m
Input signals

• **Some general comments**
  • Choice of the input signal affects significantly the quality of the estimated model
  • The estimate of a system improves for “richer” input signals (ideally – white noise, p.e. of any order)
  • The estimates are more accurate around the frequency range where the input signal has more energy
    • For instance, if we want to estimate a system with a low-pass characteristic, we may design the input signal accordingly, e.g. use low-pass filtering to pronounce the corresponding desired frequencies more by using standard filter design, increasing the clock period or decreasing the probability of level change when implementing PRBS (Soderstrom – 5.3).
    • Generally, the input signal should excite the frequencies of the system response that are of interest to us
  • Often, there are limitations as to the input signals that we can implement experimentally
The least squares estimate for LTI systems

- Return to the LS estimate for LTI systems
- We saw that for a model of the form

\[ y(t) + a_1 y(t-1) + \ldots + a_{n_a} y(t-n_a) = b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b) + e(t) \Rightarrow A(q^{-1})y(t) = B(q^{-1})u(t) + e(t) \]

\[ y(t) = \phi^T(t)\theta \]

\[ \theta = [a_1 \quad \ldots \quad a_{n_a} \quad b_1 \quad \ldots \quad b_{n_b}]^T \]

\[ \phi(t) = [-y(t-1) \quad \ldots \quad -y(t-n_a) \quad u(t-1) \quad \ldots \quad u(t-n_b)]^T \]

the least squares solution is:

\[ \hat{\theta}_{LS} = (\Phi^T\Phi)^{-1}\Phi^Ty \quad \hat{\theta}_{LS} = \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)\Phi^T(t) \right]^{-1} \frac{1}{N} \sum_{t=1}^{N} \phi(t)y(t) \] (1)

- The statistical properties of the estimates depend on the structure of the vector \( \phi(t) \). We have the following two basic cases:
  - The vector \( \phi(t) \) is deterministic (it contains present/previous values of \( u(t) \) only)
  - The vector \( \phi(t) \) is stochastic (it contains previous values of \( y(t) \) as well)
The least squares estimate for LTI systems

- Assuming that the true system is:
  \[ A_0(q^{-1})y(t) = B_0(q^{-1})u(t) + e_0(t) \]
  \[ y(t) = \phi^T(t)\theta_0 + e_0(t) \]

- We have:
  \[ \hat{\theta}_N - \theta_0 = \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)\phi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)y(t) - \theta_0 \right] = \]
  \[ = \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)\phi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)[\phi^T(t)\theta_0 + e_0(t)] - \theta_0 \right] = \]
  \[ = \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)\phi^T(t) \right]^{-1} \left[ \frac{1}{N} \sum_{t=1}^{N} \phi(t)e_0(t) \right] \]

- For \( N \to \infty \) the sums above tend to the corresponding expected values \( E\{\} \)
  - Therefore, in order to have an unbiased estimate of \( \theta_0 \) the following should hold:
    - The matrix \( E\{\phi(t)\phi^T(t)\} \) should be non-singular
    - The properties of this matrix depend on the input signal characteristics \( u(t) \). This holds for persistently exciting inputs.
  - \( E\{\phi(t)e_0(t)\} = 0 \). This holds when:
    - \( e(t) \) is white noise
    - \( e(t) \) is zero-mean but non-white, the input \( u(t) \) is uncorrelated to the noise and there are no terms of the form \( y(t-n) \). If there are such terms, the vector \( \phi(t) \) e.g. contains \( y(t-1) \), which depends on \( e(t-1) \), which in turn is correlated to \( e(t) \). This condition often does not hold.
Linear system parametrizations

• Consider the general model shown in the figure:

\[ y(t) = \sum_{\tau=1}^{\infty} g(\tau) u(t-\tau) + \sum_{\tau=0}^{\infty} h(\tau) e(t-\tau) = G(q^{-1}) u(t) + H(q^{-1}) e(t), \quad h(0) = 1 \]

\[ G(q^{-1}) = \sum_{\tau=1}^{\infty} g(\tau) q^{-\tau}, \quad H(q^{-1}) = 1 + \sum_{\tau=1}^{\infty} h(\tau) q^{-\tau} \]

• In order to fully define such a model, we need the pdf of the noise process \( e(t) \)

• Typically, we assume that \( e(t) \) is Gaussian white noise, therefore its properties are fully defined by the first and second moments (mean value and variance). In this case the spectrum of \( u(t) \) is:

\[ \Phi_u(\omega) = \lambda^2 |H(\omega)|^2 \]

• In practice, we express the transfer functions \( G,H \) as a function of a finite number of parameters either using rational transfer functions \( A(q^{-1}) \) and \( B(q^{-1}) \) or finite order state-space models

• As it is difficult to specify the exact form of \( G,H \) a priori, we use this vector of parameters \( \theta \), in other words we parametrize the system and we estimate \( \theta \) from our input/output observations:

\[ y(t) = G(q^{-1}, \theta) u(t) + H(q^{-1}, \theta) e(t), \quad E\{ee^T\} = \text{diag}(\Lambda(\theta)) \]

• The parameter vector \( \theta \), of dimension \( d \), assumes value in a subset of \( \mathbb{R}^d \):

\[ \theta \in D_M \subset \mathbb{R}^d \]

• This subset is typically selected such that \( H^{-1}(q^{-1}, \theta) \) and \( H^{-1}(q^{-1}, \theta) G(q^{-1}, \theta) \) are asymptotically stable and \( \Lambda(\theta) \) is non-negative definite
Linear system parametrizations

\[ y(t) = G(q, \Theta)u(t) + H(q, \Theta)e(t), \quad E\{ee^T\} = \text{diag}(\lambda_i^2(\Theta)) \]
\[ \Theta \in D_M \subset \mathbb{R}^d \]

- These equations define a **model set** of which the most suitable should be selected.
- A common parametrization procedure is to express \(G(q^{-1})\), \(H(q^{-1})\) as rational functions of \(q^{-1}\).
- In the general case:

\[
A(q^{-1})y(t) = B(q^{-1})u(t) + C(q^{-1})e(t)
\]

\[
G(q^{-1}, \Theta) = \frac{B(q^{-1})}{A(q^{-1})} = \frac{b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b}}{1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}}
\]

\[
H(q^{-1}, \Theta) = \frac{C(q^{-1})}{A(q^{-1})} = \frac{1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}}{1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a}}
\]

\[ \Theta = [a_1 \ldots a_{n_a} b_1 \ldots b_{n_b} c_1 \ldots c_{n_c}]^T \]

- **ARMAX model** (AutoRegressive Moving Average with eXogenous input)
- We can also include the (unknown) noise variance in the parameter vector, i.e.: \( \Theta' = [\Theta^T \ \lambda^2]^T \)
- For this formulation, the model set is given by \( D = [\Theta \mid C(z) \text{ has all zeros inside the unit circle}]^T \)
Linear system parametrizations

• Special cases:
  • $n_b=n_c=0$ – Autoregressive model (AR) (μοντέλο αυτοπαλινδρόμησης)
    \[ A(q^{-1})y(t) = e(t) \]
    \[ \theta = [a_1 \ldots a_{n_a}]^T \]
  • $n_a=n_b=0$ – Moving average model (MA) (μοντέλο κινητού μέσου)
    \[ y(t) = C(q^{-1})e(t) \]
    \[ \theta = [c_1 \ldots c_{n_c}]^T \]
  • $n_b=0$ – Autoregressive moving average model (ARMA)
    \[ A(q^{-1})y(t) = C(q^{-1})e(t) \]
    \[ \theta = [a_1 \ldots a_{n_a} c_1 \ldots c_{n_c}]^T \]

➢ In these cases we just model the time series $y(t)$ (time-series modeling) – we do not have an external input!
Linear system parametrizations

- \( n_c=0 \) – ARX model

\[
y(t) + a_1 y(t-1) + ... + a_{n_a} y(t-n_a) = b_1 u(t-1) + ... + b_{n_b} u(t-n_b) + e(t)
\]

\[
\theta = [a_1 \ldots a_{n_a} b_1 \ldots b_{n_b}]^T
\]

\[
y(t) = \frac{B(q^{-1})}{A(q^{-1})} u(t) + \frac{1}{A(q^{-1})} e(t)
\]

- Less flexibility in noise modeling
- The noise process \( v(t) \) is modeled as an AR process
- Widely used models
- We can view this model as a linear regression problem as mentioned before
  - The independent variables are not deterministic (there are terms of the form \( y(t-n) \))

- \( n_a=n_c=0 \) – Finite impulse response (FIR) model

\[
y(t) = b_1 u(t-1) + ... + b_{n_b} u(t-n_b) + e(t)
\]

\[
\theta = [b_1 \ldots b_{n_b}]^T
\]

\[
y(t) = B(q^{-1}) u(t) + e(t)
\]

- We do not have previous output values in our model – equivalent to impulse response model
- Linear regression formulation
  - The independent variables are deterministic – if consider the input as deterministic (there are no terms of the form \( y(t-n) \))
- Typically needs the estimation of more free parameters compared to AR/ARMAX models
Linear system parametrizations

• Two alternative forms of the ARMAX model
  • We can model the noise as an AR process
    \[ A(q^{-1})y(t) = B(q^{-1})u(t) + \frac{1}{D(q^{-1})} e(t) \]
    \[ D(q^{-1}) = d_1 q^{-1} + \ldots + d_{n_d} q^{-n_d} \]
  • or more generally as an ARMA process, i.e.:
    \[ A(q^{-1})y(t) = B(q^{-1})u(t) + \frac{C(q^{-1})}{D(q^{-1})} e(t) \]

• In all the previous cases, the transfer functions \( G(q^{-1}) \) and \( H(q^{-1}) \) have \( A(q^{-1}) \) as a common factor
  • It is reasonable to assume that it is more realistic to parametrize \( G, H \) independently of each other, i.e.:
    \[ y(t) = \frac{B(q^{-1})}{F(q^{-1})} u(t) + \frac{C(q^{-1})}{D(q^{-1})} e(t) \]
Linear system parametrizations

• In the simplest case, we have:
  \[ w(t) + f_1 w(t-1) + \ldots + f_{n_f} w(t-n_f) = b_1 u(t-1) + \ldots + b_{n_b} u(t-n_b) \]
  \[ y(t) = w(t) + e(t) \]
  \[ \theta = [b_1 \ldots b_{n_b} f_1 \ldots f_{n_f}]^T \]
  \[ e(t) = y(t) - \frac{B(q^{-1})}{F(q^{-1})} u(t) \]

• Therefore, there are several options for the parametrization of an LTI system.
  • All these may be viewed as special cases of the general model:
  \[ y(t) = \sum_{\tau=1}^{\infty} g(\tau) u(t-\tau) + \sum_{\tau=0}^{\infty} h(\tau) e(t-\tau) = G(q^{-1}) u(t) + H(q^{-1}) e(t) \]

• In some of them we include previous output values
  • Less free parameters than FIR models
• The output noise is incorporated in different ways
  • Some model types (e.g. Box-Jenkins) offer greater flexibility but at the cost of increased complexity (more parameters)
• In the next lectures we will examine how the model parameters may be estimated